

## 7.1 Compassion Is Not Always a Motivated Choice: A Multiple Decision Systems Perspective

M. A. Montgomery, Andreas Kappes, and M. J. Crockett

Cameron and Rapier (hereinafter CR) propose that compassion—defined as encompassing both the motivation to help others and the resulting behavior—is a conscious choice. People are capable of evaluating possible outcomes of their actions and making decisions that maximize desired outcomes, and this holds true for compassion. CR highlight research that shows our ability to manage the competing costs and benefits of compassion and to choose self-interested or altruistic actions accordingly. Here we suggest that compassion is not always a deliberative choice “to prioritize and balance competing goals within specific contexts” (CR, this volume). Rather, the extent to which compassion is a choice may depend on familiarity with the action itself and the target of compassion. We may deliberately choose whether or not to have compassion for faraway strangers, but compassion toward kin, friends, and in-group members often feels inevitable. Recent advances in the neuroscience of decision-making shed light on this critical distinction.

Choosing to feel compassion for others and act accordingly, like other kinds of value-based choices, reflects the input of multiple decision systems that are psychologically and neurally dissociable (Crockett, 2013; Dolan and Dayan, 2013; Gešiarz and Crockett, 2015). The *goal-directed system* selects actions that will bring about desirable end states. The *habitual system* selects actions based on their reinforcement history. While the goal-directed system dominates in unfamiliar situations with novel stimuli, the habitual system is more likely to take the reins when dealing with familiar situations (Daw, Niv, and Dayan, 2005). A third, *Pavlovian system* responds instinctively to ecologically significant stimuli. Most of the studies marshalled to support CR’s thesis investigated compassion for strangers in distant countries—precisely the kinds of situations that would rely disproportionately on the goal-directed system. Unsurprisingly, then, compassion in such situations resembles a deliberative, goal-directed choice. However,



we argue that compassion for more familiar targets may not always result from “motivated choices to prioritize and balance competing goals” (CR, this volume). When directed toward our nearest and dearest, compassion may more closely resemble a habit or an instinct.

While the values of the goal-directed system are based on current goals, the values of the habitual system reflect past goals which might run counter to current ones. The habitual system bases decisions on past experience, automatically selecting actions that have produced good outcomes in the past. In this sense, it chooses based on the value of the actions themselves rather than the consequences those actions produce—for example, a rat trained to press a lever for food will come to assign value to the pressing of the lever. Unlike the goal-directed system, the habitual system is highly efficient but is unable to adjust action selection on the basis of changes in the value of outcomes or motivation to obtain those outcomes (Balleine and O’Doherty, 2009). Habitual action selection becomes apparent in devaluation procedures: Rats trained extensively to press a lever for food will continue to press the lever even when the food is rendered less valuable—for instance, if the rat is sated. In contrast, the goal-directed system is sensitive to changes in the value of outcomes—so if the goal-directed system is in control, the rat will stop pressing the lever when sated (Balleine and O’Doherty, 2009).

There is evidence that in humans, compassion can resemble a habit. That is, compassionate choices may sometimes be implemented without thought of their consequences, just like the sated rat that continues to press the food-bearing lever. In other words, if compassionate acts are frequently rewarded, the act itself may come to take on intrinsic value. Once this occurs, subsequent compassionate acts may occur rather mindlessly, without paying heed to the consequences of those acts. This notion is captured by the “warm glow” theory of altruism, which posits that people behave generously simply because it feels good, irrespective of the consequences of giving (Andreoni, 1990). In one exemplary study, participants chose their favorite charity and then had the chance to donate a portion of a \$10 endowment to the charity (Crumpler and Grossman, 2008). However, the chosen charity also received a \$10 endowment, and for each dollar donated by the participant, the experimenter deducted a dollar from the charity’s endowment. Thus, the charity would always end up with \$10, regardless of whether the participants donated all or none of their endowment. Nearly 60% of participants chose to donate, despite the fact that donating had no effect on the outcome of the charity. This suggests that participants did not donate on the basis of a deliberate choice to maximize the welfare of the



charity. Rather, donation resembled a habit, in the sense that it was insensitive to the consequences of donating for the charity. Instead they likely donated because it simply felt good. Indeed, such “noncontingent helping” is a core feature of many close relationships where people help each other without keeping track of whether the favor will be returned (Clark and Mills, 2012). Other work has demonstrated the establishment of habit-like prosocial behavior in the laboratory. Participants who played a prisoner’s dilemma in an environment where the payoff structure promoted cooperation were more likely to cooperate in subsequent tests, and were faster to do so, suggesting repeated reinforcement of cooperative behavior can create “habits of virtue” (Peysakhovich and Rand, 2013; see also Rand et al., 2014). Thus, repeated positive experience with compassionate acts may create circumstances in which compassion is no longer a deliberate choice, but rather a habit that is not based on a cost–benefit analysis of competing goals, as CR suggest.

In contrast to the habitual and goal-directed systems, the Pavlovian system automatically deploys built-in solutions (Dayan and Berridge, 2014; Robinson and Berridge, 2013; Wiering and van Otterlo, 2012). Without the need for past experience, these behaviors occur reflexively whenever a given situation arises and are based on classes of stimuli that were important in our evolutionary past. Unlike the habitual and goal-directed systems, which can in principle learn to emit any kind of action, the Pavlovian system is constrained to a limited set of innate responses to stimuli (e.g., approach stimuli associated with rewards, avoid stimuli associated with punishments). In other words, the Pavlovian system deploys actions that have been successful in our collective, evolutionary history, while the habitual system deploys actions that have been successful in our individual history. Although these two systems differ, both are susceptible to making suboptimal recommendations in atypical contexts since they rely on retrospective rather than prospective data (Carmel and Markovitch, 1998; Killeen, 2003; Williams and Williams, 1969).

Evidence of compassionate behavior in very young children suggests there may be an instinctual (i.e., a Pavlovian) component to compassion. Children as young as 18 months help others to achieve their goals (Warneken, 2013; Warneken and Tomasello, 2006). Eighteen-month-olds, for instance, saw an experimenter trying to put a stack of magazines into a cabinet which he could not open since his hands were full. The children spontaneously got up and opened the door for him (Warneken and Tomasello, 2006). These infants cannot talk yet, let alone make any calculation of predicted outcome value (Klossek, Russell, and Dickinson, 2008;

Kenward, Folke, Holmberg, Johansson, and Gredebäck, 2009). It therefore seems unlikely that a deliberation process guides such actions, ruling out a role for the goal-directed system. Perhaps parents may have encouraged and rewarded prosocial behaviors during the short life of the infant, raising the possibility that such prosocial actions could be habitual. But it might be fair to assume that an 18-month-old will not have had many opportunities to help others or even be urged to help; thus, a role for the habitual system also seems unlikely. Furthermore, rewards are not needed to instigate prosocial behavior in infants and might even undermine prosocial behavior (Warneken and Tomasello, 2014). These findings suggest that humans may be born with instincts for acting compassionately, potentially independent of culture and socialization (House et al., 2013; Romero, Castellanos, and de Waal, 2010).

The empathy–altruism hypothesis (Batson, Duncan, Ackerman, Buckley, and Birch, 1981) suggests that instincts for helping in young children may be automatically triggered by facial expressions of sadness and suffering. Around 15 months, infants are able to understand that their emotions automatically match those of another person, and that they originate from the state of that other person (Bischof-Köhler, 2012). Moreover, empathy can induce helping even in situations where helping leads to maladaptive outcomes for oneself or for others in greater need (Batson, Batson, et al., 1995; Batson, Klein, Highberger, and Shaw, 1995). Charles T. Carbonell Sr., for instance, is the only living person to have twice received the Carnegie Award, which recognizes outstanding acts of altruism. Carbonell rescued a police officer in a fight and a woman from inside a burning car. When asked about why he did these heroic acts, he answered, “If I’d thought about it first, I probably wouldn’t have done it. I always do it first and then think after” (Majors, 2014). A recent quantitative survey of Carnegie Award winners revealed that the majority of them acted quickly and intuitively, rather than deliberating (Rand and Epstein, 2014). Instinctive compassionate drives are not governed by deliberate choices and might even run counter to them: Four of the other award winners that year died while trying to save others from burning, drowning, or assault (Rand and Epstein, 2014). Furthermore, given the extraordinary circumstances of these situations, it is unlikely that helping in these situations is triggered by habits based on past experience. Rather, the instinctive empathic reaction to the face of somebody suffering might have triggered the compassionate behavior.

Instinctive compassion also puts limits on the regulation of compassion. Pavlovian reflexes for compassion might be tied to our relationships with kin (De Waal, 2008; Preston and De Waal, 2002), suggesting limited control

over compassion when it involves our children's safety and well-being: It seems rather counterintuitive to think that parents would deliberate about whether to feel compassion for their children. By extension, Pavlovian reflexes may explain increased compassion for in-group members relative to out-group members. From infancy human beings prefer familiar and same-race faces to novel face stimuli (Barrile, Armstrong, and Bower, 1999; Kelly et al., 2005) and show enhanced compassion toward partners who have similar facial features to themselves (Krupp, DeBruine, and Barclay, 2008) or who belong to the same group (Ahmed, 2007; Dunham, Baron, and Carey, 2011).

CR suggest that callousness toward out-group members can also be explained as a motivated choice: People expect the costs of helping out-group members or stigmatized people to be especially high and, consequently, refrain from helping. We suggest an alternative (or at least an additional) explanation for these findings. Perceiving distress in the faces of others is one of the most important cues for instinctive compassion (Goetz, Keltner, and Simon-Thomas, 2010). Hence, a Pavlovian explanation for callousness toward out-group members would suggest that people may have trouble perceiving distress in the faces of out-group members, which reduces the likelihood of triggering an instinctive compassion response. In other words, it may not be the case that reduced compassion toward out-group members is the result of a motivated self-regulation process, as CR suggest, but rather the absence of an instinctive compassion response that requires perception of distress cues in order to be triggered. It is well-known that humans and even other primates are less skillful at processing the faces of out-group members (Dahl, Rasch, and Chen, 2014). Importantly, this other-race effect leads to reduced empathy when observing an out-group member in pain, both at the behavioral and neural level (Avenanti, Sirigu, and Aglioti, 2010; Xu, Zuo, Wang, and Han, 2009). Importantly, a reduced neural response to pain in out-group members predicts reduced helping of out-group members (Hein, Silani, Preuschhoff, Batson, and Singer, 2010). These findings do not mean we are constitutionally incapable of showing empathy and compassion toward out-group members; in fact, increased contact with out-group members not only decreases the other-race effect in pain perception (Cao, Contreras-Huerta, McFadyen, and Cunningham, 2015) but also increases compassionate behavior toward out-group members (Brown and Hewstone, 2005). This suggests that once people can reliably detect pain in the faces of out-group members, the same Pavlovian instincts may govern their compassion for in-group and out-group members.

We are not claiming that the habitual or Pavlovian systems are more influential in driving compassionate behavior than the goal-directed system; in many situations all three systems work together to guide compassion. For instance, during the development of habitual helping, the goal-directed system is crucial. Wanting to be a well-respected member of the community will frequently lead to the repeated, rewarding implementation of helping behavior, which is necessary for the formation of habitual helping. Nevertheless, we suggest that only considering deliberative aspects of compassion leads to an incomplete account of compassion. The control we have over our compassion has limits. These limits are tied to our collective evolutionary and individual history; express themselves in automatic tendencies to help some people, but not others; and emerge when the person in need is not our child but a distant stranger.

## References

- Ahmed, A. M. (2007). Group identity, social distance and intergroup bias. *Journal of Economic Psychology*, 28(3), 324–337.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal*, 100(401), 464–477. doi:10.2307/2234133
- Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology*, 20(11), 1018–1022. doi:10.1016/j.cub.2010.03.071
- Balleine, B. W., & O'Doherty, J. P. (2009). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, 35(1), 48–69. doi:10.1038/npp.2009.131
- Barrile, M., Armstrong, E. S., & Bower, T. G. R. (1999). Novelty and frequency as determinants of newborn preference. *Developmental Science*, 2(1), 47–52.
- Batson, C. D., Batson, J. G., Todd, R. M., Brummett, B. H., Shaw, L. L., & Aldeguer, C. M. (1995). Empathy and the collective good: Caring for one of the others in a social dilemma. *Journal of Personality and Social Psychology*, 68(4), 619–631.
- Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, 40(2), 290–302. doi:10.1037/0022-3514.40.2.290
- Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology*, 68(6), 1042–1054

- Bischof-Köhler, D. (2012). Empathy and self-recognition in phylogenetic and ontogenetic perspective. *Emotion Review*, 4(1), 40–48. doi:10.1177/1754073911421377
- Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 255–343). San Diego, CA: Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0065260105370055>.
- Cao, Y., Contreras-Huerta, L. S., McFadyen, J., & Cunnington, R. (2015). Racial bias in neural response to others' pain is reduced with other-race contact. *Cortex*, 70, 68–78. doi:10.1016/j.cortex.2015.02.010
- Carmel, D., & Markovitch, S. (1998). Model-based learning of interaction strategies in multi-agent systems. *Journal of Experimental & Theoretical Artificial Intelligence*, 10(3), 309–332.
- Clark, M. S., & Mills, J. R. (2012). A theory of communal (and exchange) relationships. *Handbook of Theories of Social Psychology*, 2, 232–250.
- Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, 17(8), 363–366. doi:10.1016/j.tics.2013.06.005
- Crumpler, H., & Grossman, P. (2008). An experimental test of warm glow giving. *Journal of Public Economics*, 92, 1011–1021.
- Dahl, C. D., Rasch, M. J., & Chen, C.-C. (2014). The other-race and other-species effects in face perception—A subordinate-level analysis. *Frontiers in Psychology*, 5(1068). doi:10.3389/fpsyg.2014.01068
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711. doi:10.1038/nn1560
- Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective & Behavioral Neuroscience*, 14(2), 473–492. doi:10.3758/s13415-014-0277-8.
- De Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, 59, 279–300.
- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2), 312–325. doi:10.1016/j.neuron.2013.09.007
- Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of “minimal” group affiliations in children. *Child Development*, 82(3), 793–811. doi:10.1111/j.1467-8624.2011.01577.x
- Gęsziarz, F., & Crockett, M. J. (2015). Goal-directed, habitual and Pavlovian prosocial behavior. *Frontiers in Behavioral Neuroscience*, 9(135).

Goetz, J. L., Keltner, D., & Simon-Thomas, E. (2010). Compassion: An evolutionary analysis and empirical review. *Psychological Bulletin*, 136(3), 351–374. doi:10.1037/a0018807

Hein, G., Silani, G., Preuschhoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68(1), 149–160. doi:10.1016/j.neuron.2010.09.003

House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., & Boyette, A. H., et al. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, 110(36), 14586–14591.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., et al. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, 8(6), F31–F36.

Kenward, B., Folke, S., Holmberg, J., Johansson, A., & Gredebäck, G. (2009). Goal directedness and decision making in infants. *Developmental Psychology*, 45(3), 809–819. doi:10.1037/a0014076

Killeen, P. R. (2003). Complex dynamic processes in sign tracking with an omission contingency (negative automaintenance). *Journal of Experimental Psychology. Animal Behavior Processes*, 29(1), 49–61.

Klossek, U. M., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology: General*, 137(1), 39–51. Retrieved from <http://doi.org/10.1037/0096-3445.137.1.39>

Krupp, D. B., Debruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior*, 29(1), 49–55.

Majors, D. (2014). Florida man wins Carnegie Award for heroism a second time. *Pittsburgh Post-Gazette*, December 17. Retrieved November 26, 2015, from <http://www.post-gazette.com/local/city/2014/12/17/Florida-man-wins-Carnegie-Award-for-heroism-a-second-time/stories/201412170089>

Peysakhovich, A., & Rand, D. G. (2013). Habits of virtue: creating norms of cooperation and defection in the laboratory. Available at SSRN 2294242. Retrieved from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2294242](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2294242)

Preston, S. D., & De Waal, F. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(01), 1–20.

Rand, D. G., & Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLoS One*, 9(10), e109687. doi:10.1371/journal.pone.0109687



Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., et al. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5. <http://doi.org/10.1038/ncomms4677>.

Robinson, M. J., & Berridge, K. C. (2013). Instant transformation of learned repulsion into motivational "wanting." *Current Biology*, 23, 282–289.

Romero, T., Castellanos, M. A., & de Waal, F. B. (2010). Consolation as possible expression of sympathetic concern among chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, 107(27), 12110–12115.

Wiering, M., & van Otterlo, M. (Eds.). (2012). *Reinforcement learning* (Vol. 12). Berlin, Germany: Springer. Retrieved from <http://link.springer.com/10.1007/978-3-642-27645-3>.

Warneken, F. (2013). The origins of human cooperation from a developmental and comparative perspective. *Evolution of Mind, Brain, and Culture*, 5, 149.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311(5765), 1301–1303.

Warneken, F., & Tomasello, M. (2014). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Motivation Science*, 1(S), 43–48. <http://doi.org/10.1037/2333-8113.1.S.43>

Williams, D. R., & Williams, H. (1969). Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior*, 12(4), 511–520.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, 29(26), 8525–8529. doi:10.1523/JNEUROSCI.2418-09.2009

**PROPERTY OF THE MIT PRESS  
FOR PROOFREADING, INDEXING, AND PROMOTIONAL PURPOSES ONLY**